REGIONAL FLOOD FREQUENCY ANALYSIS ACCOUNTING FOR SPORADIC THUNDERSTORMS IN NORTH CENTRAL OREGON

Henry Hu, Senior Hydraulic Engineer, WEST Consultants, Inc., 12509 Bel-Red Road, Suite 100, Bellevue, WA 98005, hhu@westconsultants.com; Todd Bennett, Senior Project Engineer, WEST Consultants, Inc., Bellevue, WA, tbennett@westconsultants.com; Wilbert Thomas, Jr., Senior Technical Consultant, Michael Baker, Jr., Inc., Alexandria, VA, wthomas@mbakercorp.com; Joseph Weber, Federal Emergency Management Agency, Region X, Bothell, WA, joseph.weber@dhs.gov.

Abstract: Flood-frequency relations for streams in the Willow Creek basin and other small streams in north central Oregon are difficult to estimate, largely because of the extreme temporal and spatial variability of thunderstorm-driven floods. In this region, summer thunderstorms typically occur over small, isolated, and ungaged basins with excessive amounts of rainfall within a very short time period, leading to extreme flood peaks. Due to the limitations of a conventional flood frequency regression method, the majority of flood events from large thunderstorms are unrepresented in the systematic flow record used to develop conventional regional regression equations. This can significantly underestimate frequency flows in this region, especially for smaller basins where thunderstorms have more impacts. This study applies a hybrid regression method to develop a set of regional regression equations by incorporating additional thunderstorm summer peaks either at isolated and ungaged sites or at gaging stations with short records. The base flood discharges for Willow Creek and its tributaries are then estimated using the regional hybrid regression equation and are compared with gaging data and other frequency flows. The comparison indicates that the hybrid regression results are reasonable.

INTRODUCTION

North central Oregon is a semi-arid region that experiences many severe and destructive floods with a majority of these being caused by summer thunderstorm events. Summer thunderstorms, although sporadic, can generate excessive amounts of rainfall within a very short time period, leading to extreme flood peaks. This can have a significant impact on smaller watersheds, where the peak discharge per square mile is often among the highest recorded in the United States. In this region, summer thunderstorms typically occur over small, isolated, and ungaged basins. Although some gaging stations may record large summer peaks, the data records are usually short with many years of zero flows due to the nature of intermittent or ephemeral streams in this semi-arid region. This leads to the majority of flood events from large thunderstorms being unrepresented in the systematic flow record and, therefore, also in the data used to develop conventional regional regression equations. This can significantly underestimate calculated frequency flows in this region, especially for smaller basins where thunderstorms have more impacts.

This study applies a hybrid regression method originally proposed by Hjalmarson and Thomas (1992) to develop a set of regional regression equations that incorporate isolated thunderstorm summer peaks. A distinguishing characteristic of the hybrid method is that it combines all peak flow records within a hydrologically similar region into one data set. This includes all annual peaks (zero or non-zero flows) at gaging stations with long-term or short-term records and a number of historic thunderstorm peak discharges at ungaged sites. Another advantage of the hybrid method is that it does not require the assumption of a theoretical probability distribution, if a plotting-position formula is used, since in semi-arid and arid regions many station flood frequency relations are typically undefined or unreliable if fitted with a theoretical distribution.

In this paper, the basic steps for performing a hybrid regression analysis are first described. Then, one test case for evaluating the applicability and accuracy of the hybrid method in the study region is presented. Finally, the hybrid method is applied to Willow Creek and its tributaries in Morrow County, Oregon to develop a set of regression equations for use in a Federal Emergency Management Agency (FEMA) Flood Insurance Study. The 1% annual chance flood discharges estimated from the hybrid regression method for Willow Creek and its tributaries are evaluated and compared with other available hydrologic data.

HYBRID REGRESSION METHOD

The hybrid method for a regional regression analysis was described in detail by Hjalmarson and Thomas (1992). It is based on the station-year method (Fuller 1914) of a frequency analysis to produce regional flood-frequency relations. The station-year method is based on the assumption that independent records of annual peak discharge from a region can be combined to form one long composite record if the peaks of the individual records can be reduced to a common base. Spatial sampling is assumed to be equivalent to time sampling if the records are reasonably independent. Therefore, for example, a combination of 10 gaging stations, each with 10 years of records, results in a 100-year composite record.

The hybrid method starts with forming a single data set by pooling annual peaks from many gaging stations and historic flood estimates at ungaged sites with the assumption that the annual peaks are reasonably independent. It uses the following regression equation that is commonly used in many regional flood-frequency analyses:

$$Q = aA^{b}B^{c}C^{d} \qquad (1)$$

where Q is the discharge for different exceedance probabilities; a is the coefficient; A, B, C are independent basin and climatic parameters; and b, c, d are regression exponents. Because drainage area is the most significant independent variable that affects flood characteristics, the hybrid method starts the regression between discharge and drainage area. It involves the following steps:

Step 1. The drainage area for all sites is ranked from the smallest to the largest. The combined single long record is then divided into three or more groups according to basin drainage area. Each group has a number of stations (see Figure 1 for an illustration). Each station has a number of years with flow or with zero flow. To ensure an unbiased relation in the regression analysis, each group has a nearly equal sample size.

Step 2. Each peak discharge within each group is standardized by dividing by A^{b} (the exponent b is equal to one for the first iteration).

Step 3. In each group, the exceedance probabilities of the standardized peaks can be estimated either by fitting a theoretical flood-frequency curve if appropriate or simply by using a plotting-position formula. To avoid extrapolations to the 1% annual chance flood level, each group has at least 100 station-years (peaks) with flow to estimate the 0.01 probability. If an elementary plotting-position formula is used, a theoretical probability distribution is no longer required. This advantage is important because in semi-arid and arid regions, many station flood-frequency relations are typically undefined or unreliable if fitted with a theoretical curve.

In this study, the probabilities are simply computed using the Cunnane plotting-position formula (Cunnane 1978):

$$p = (m - 0.4)/(N + 0.2) \tag{2}$$

where p is the probability of a flood with rank m occurring in any given year; m is the rank number of flood values with the largest equal to 1; and N is the number of total station-years within a group. The Cunnane formula was used because it is an unbiased and relatively distribution-free plotting position, implying that it is appropriate when the underlying distribution of the data is not known.

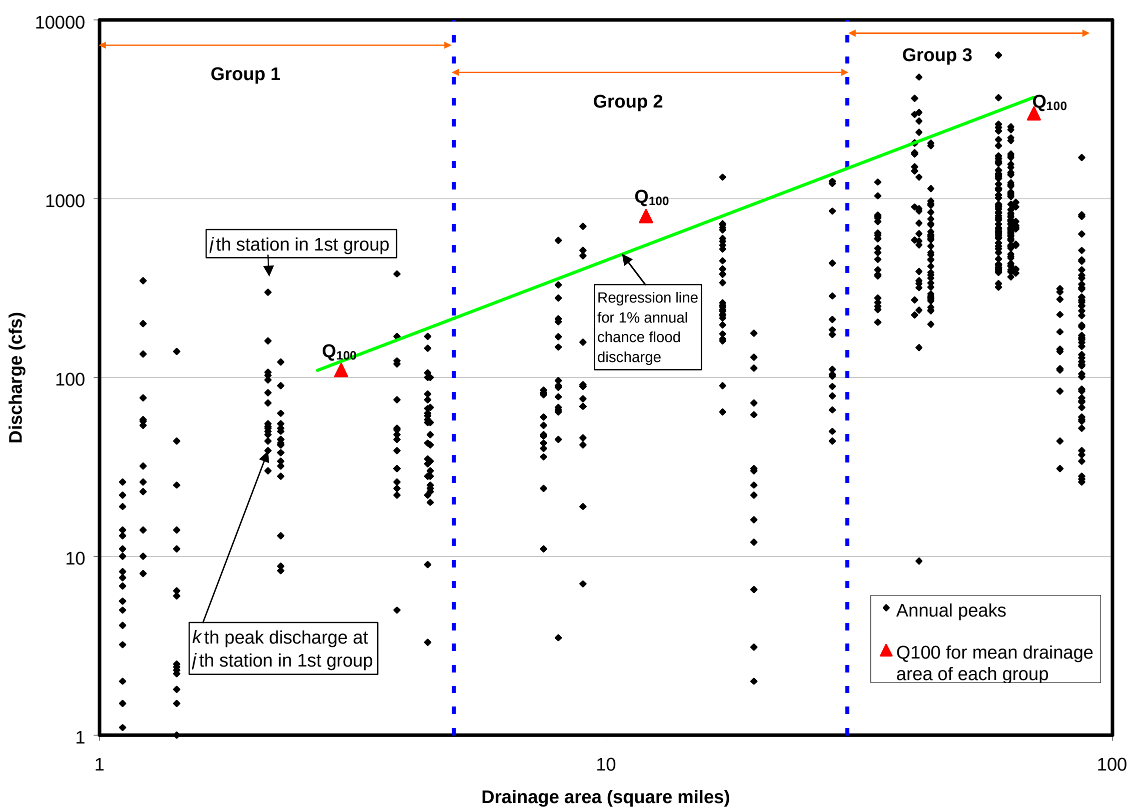

Figure 1 A sample hybrid regression relation for 1% annual chance flood discharges.

Step 4. The frequency flows for each group obtained in Step 3 are de-standardized by multiplying by the weighted geometric mean drainage area.

Step 5. For each exceedance probability, a linear regression analysis is conducted between *Q* and mean drainage area in log space (see Figure 1 for a sample regression line), and a new exponent, *b*, is computed. To perform a linear regression, the combined data set has to be divided into at least three groups.

Step 6. Using the new exponent, an iterative process that uses a regression and flood-frequency analysis is repeated until the computed exponent converges.

Each additional parameter can be separately added to the relation with the same iterative technique starting at Step 1. The new parameter (e.g., B or C) is used in place of drainage area. The original peak discharges in Step 2 are replaced with standardized discharges obtained from the last iteration for the previous parameter. The coefficient, *a*, in Equation (1) is determined during the last linear regression (in log space) of the last parameter.

TEST CASE

The hybrid method was first applied to a combined region that consists of north central Oregon defined in USGS Water Resources Investigation Report (WRIR) 82-4078 (Harris and Hubbard 1983), Washington Region 6, and the southwestern portion of Washington Region 9 defined in WRIR 97-4277 (Sumioka et al. 1998) as a test case (Figure 2). This combined region was selected because it provides a hydrologically similar region and includes all large flood peaks from thunderstorms in the region. Common features of this area are relative low elevations, lack of vegetation and forest cover, relatively uniform geology and soil composition, and high summer temperatures. The purposes for performing a test case analysis include: (1) to test the applicability of the hybrid method by comparing the regression results to Log-Pearson III frequency flows at gaging stations; (2) to compare the accuracy between the hybrid and conventional regression equations; (3) to conduct a sensitivity analysis to test how the number of groups affects the result; and (4) to test which combination of independent parameters have the better prediction ability. Because the stations are widely scattered throughout the region, and most of large peaks are caused by local thunderstorms over a small aerial extent, the combined annual peaks are assumed to be reasonably independent.

To ensure gaging stations that have similar semi-arid watershed characteristics, all stations that have annual precipitation equal to or larger than 25 inches are not included in the analysis because annual precipitation in this area is generally less than 25 inches. The remaining data set has a total of 928 station-years from 37 gaging stations. There are 25 stations in Oregon, 4 stations in Washington Region 6, and 8 stations in Washington Region 9. The same annual peak records and basin and climatic parameters as those used in the USGS studies were used in the hybrid regression analysis.

USGS WRIR 97-4277 used drainage area and annual precipitation as two independent variables to develop regression equations. An additional parameter, the mean January minimum temperature, was included in the USGS regression equations for north central Oregon. The hybrid analysis was first conducted to test which combinations of basin and climatic parameters have reasonable regression relationships. A regression relation is considered reasonable if the exponents are consistent with physical characteristics. For example, the exponent of mean annual precipitation should be a positive value in a regression equation, implying that with the increase in mean annual precipitation, runoff also increases. However, for the data set used in the test case, a combination of drainage and mean annual precipitation results in a negative exponent for mean annual precipitation when applying the hybrid method. Therefore, the mean annual precipitation should not be used in the hybrid regression equations for the selected data set. The regression analysis suggests that the mean January minimum temperature should also not be included in the regression equation in the study region.

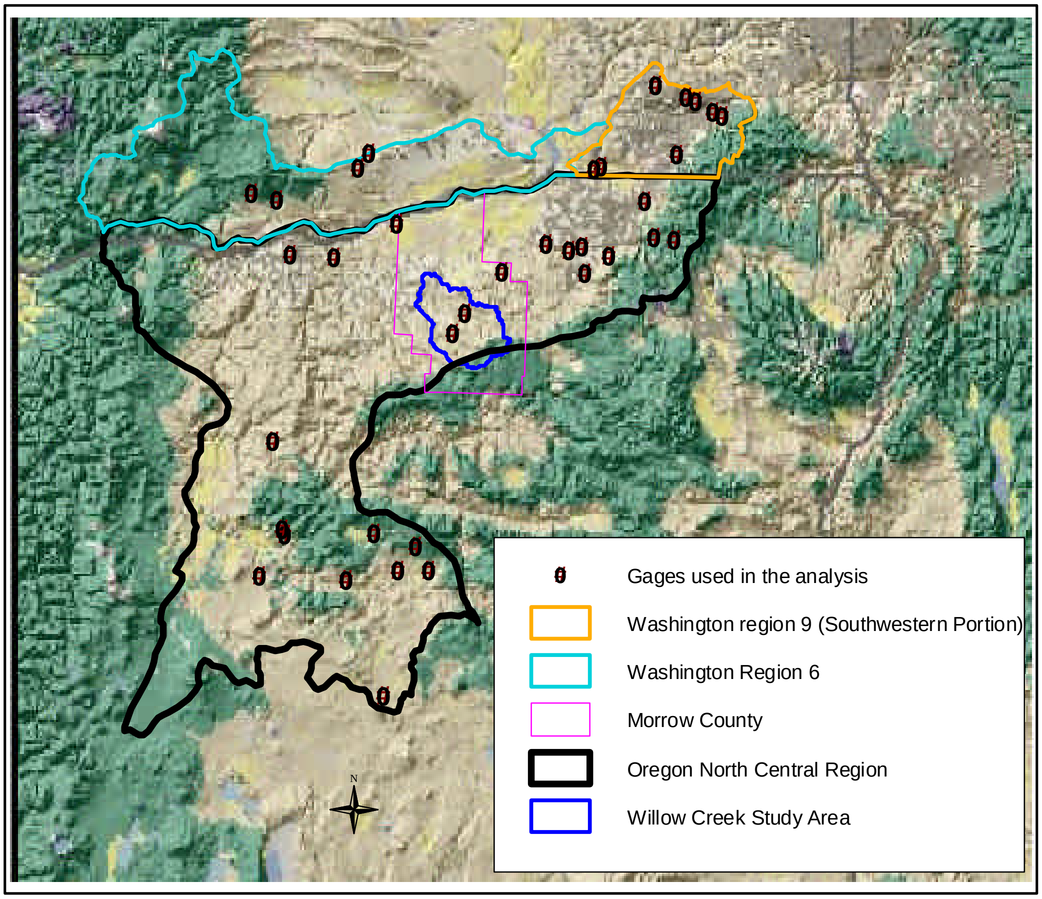

Figure 2 Distribution of gaging stations used in the test case.

For all remaining combinations that appear to have reasonable regression relations, a comparison is then made to determine which combination has the least root mean square error (RMSE). A RMSE is calculated using the hybrid regression discharges and Log-Pearson III flood frequency flows at gaging stations published in USGS WRIRs 82-4078 and 97-4277. For comparison, RMSEs were also calculated using USGS regression discharges and Log-Pearson III discharges. For gaging stations within Oregon, the Oregon regression equations were used whereas the Washington equations were used for Washington gaging stations. Figure 3 shows the variation of the RMSE versus the number of groups for three different combinations of independent variables for the 1% annual chance flood discharges. The comparison indicates that a combination of drainage area, mean basin elevation, and 50% annual chance and 6-hour precipitation intensity could yield the smallest error.

Figure 3 also shows how the number of groups affects the hybrid regression results. It is interesting to note that when the group number is close to one half of the maximum group number (8 for this test case), the regression equation could have the best prediction. In addition, the RMSEs do not change significantly when the group number is between 4 and 7. However, at the minimum number of groups (which is three), the error significantly increases. Because the linear regression (in log space) is only based on three pairs of discharges and independent variables, the regression relations may not be very reliable. If the combined record is divided into too many groups (8 for this case), the error also significantly increases. This is probably because the combined records in each group may not be as regionally representative as longer records if more gaging stations are included in each group. The comparison of RMSEs in the test case demonstrates that the hybrid regression method is a reasonable approach. With a proper combination of basin and climatic parameters and a proper group number, the hybrid method can have the same level of accuracy when compared to conventional regression equations.

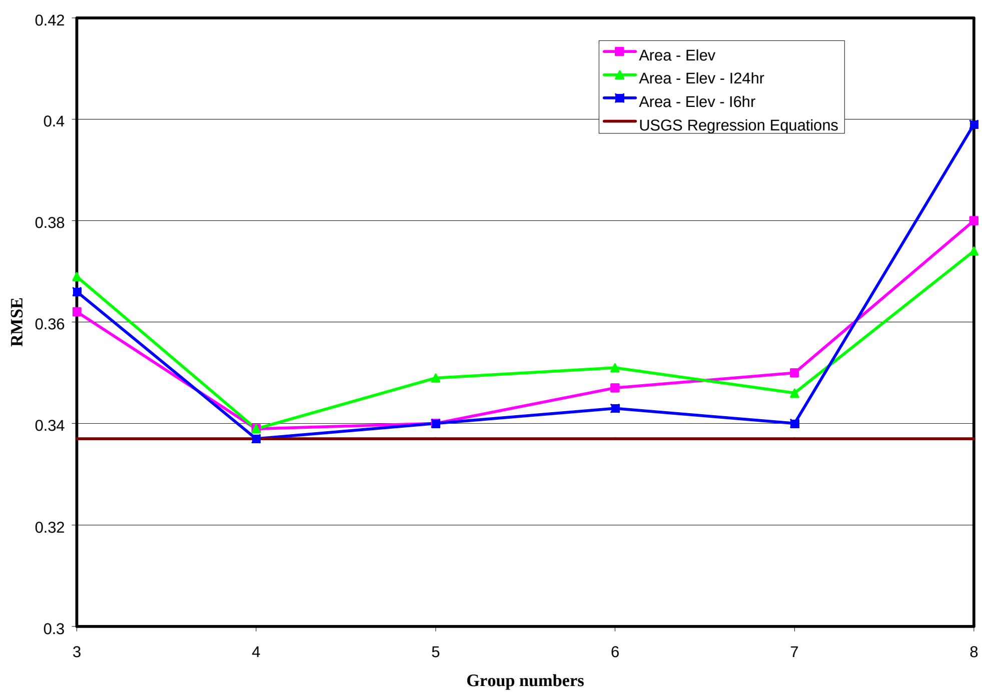

Figure 3 Relationship between group numbers and RMSEs.

APPLICATION OF HYBRID REGRESSION METHOD

The areas used in the test case were slightly modified to exclude high elevation areas just east of Cascade Mountains. The peak flow data set used in the test case was extended to include all peaks through water year 2003, many additional gaging stations with short-term records, and a few miscellaneous sites (non-gaging stations for extreme discharges at isolated locations). In the data set, there are a total of 48 peaks that have unit discharge larger than 100 cfs per square mile. Of these 48 largest peaks, 28 (19 summer peaks and 9 peaks for other seasons) came from additional gaging stations or miscellaneous sites that were not included in the data set used in USGS WRIRs 82-4078 or 97-4277. The 19 additional summer peaks reflect thunderstorms and are expected to impact the results of hybrid regression equations.

The data set contains a total of 1,401 station-years from 85 stations with 197 station-years of zero flow. The minimum number of groups is three. To avoid the extrapolation to the 1% annual chance flood discharge, the maximum number of groups is 12 (the total station-years with flow divided by 100). Based on the sensitivity of the hybrid regression equation accuracy to the number of groups as presented in the test case, the data set was divided into 6 groups (a half of the maximum number of groups).

Drainage area, mean basin elevation, and 50% annual chance and 6-hour precipitation intensity were initially used as the independent variables. However, the mean basin elevation and 50% annual chance and 6-hour precipitation intensity were not used in the final regression equation because either the regression equation itself is not reasonable or the discharges obtained for small basins in very low elevation basins were unreasonable. The regression equation for the 1% annual chance flood discharge from the final hybrid regression analysis is as follows:

$$Q_{0.01} = 10^{3.02655} A^{0.446} \tag{3}$$

where $Q_{0.01}$ is the 1% annual chance flood discharge in cfs and A is the drainage area in square miles.

EVALUATION OF FLOOD-FREQUENCY RELATIONS

Equation (3) was applied to Willow Creek and its tributaries in Morrow County for estimating the 1% annual chance flood discharges. Figure 4 shows the comparison between the proposed 1% annual chance flood discharges estimated from the hybrid regression method and other hydrologic data. It is clear that the proposed discharges are generally much smaller than the effective discharges with a few exceptions (the effective discharges are discharges used in the current FEMA flood insurance study). This concurs with prior conclusions that the effective discharges based on a rainfall-runoff model were originally overestimated. For the two smallest streams, the proposed discharges are higher than the effective discharges, resulting from the inclusion of large thunderstorm events that have occurred in the region.

The proposed 1% annual chance flood discharges are generally larger than the Oregon and Washington State gaging data. This is expected because the gaging data typically do not reflect the impact of large thunderstorms that occurred in small and ungaged basins. Table 1 shows the variation of the 1% annual chance flood discharges estimated from different approaches at gaging station 14034500 (Willow Creek at Heppner). Again, the proposed discharges using the hybrid regression method are much smaller than the unregulated FIS discharge but larger than the discharge calculated using USGS regression equations, which tend to underestimate the flow due to the absence of large summer peaks in the data, used to develop the conventional regression equations. Station 14034500 has the longest record in the Willow Creek basin and has a roughly 100 years of historic period of record, from which a Log-Pearson III frequency analysis is considered reliable. The proposed discharge is close to the Log-Pearson III frequency flow and was therefore considered reasonable.

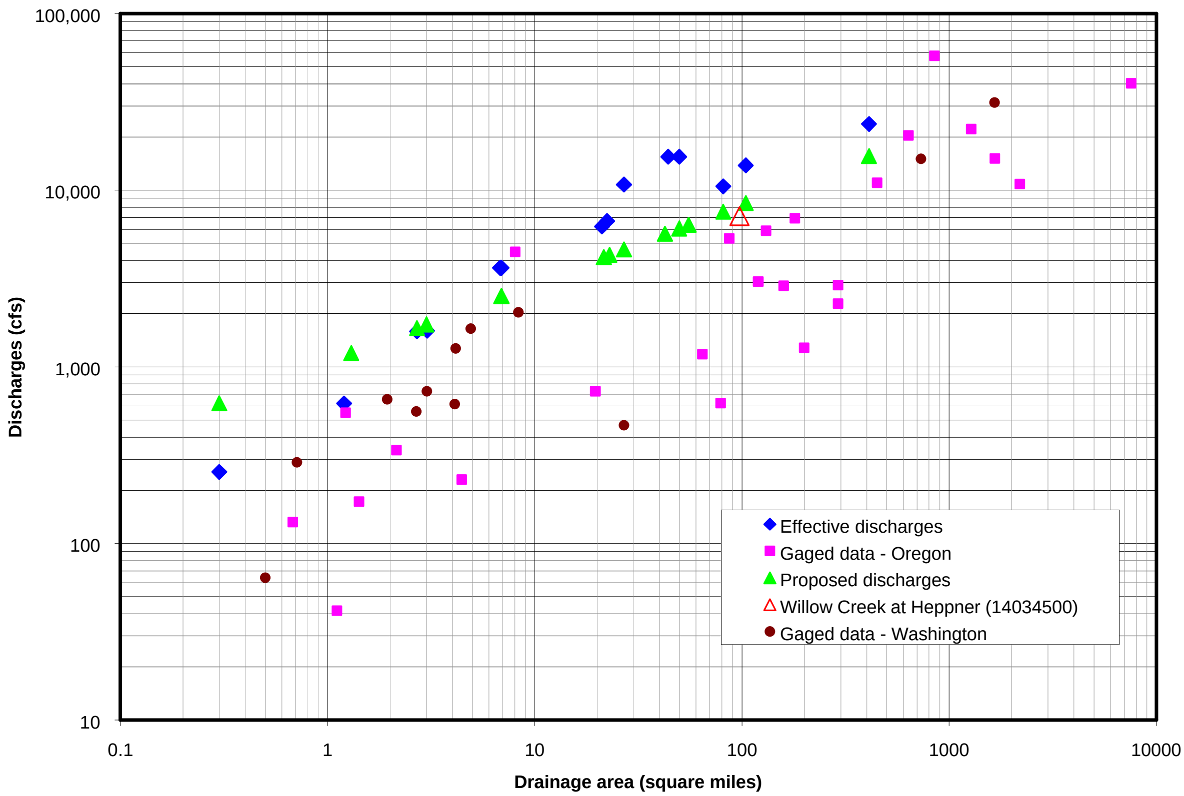

Figure 4 Comparison of 1% annual chance flood discharges.

Table 1 Comparison of 1% annual chance flood discharges at station 14034500.

Station number	Drainage area	Unregulated FIS base flood discharges	USGS eastern Oregon regression equation	USGS Washington regression equation	LP-III Frequency Flow	Hybrid regression analysis
	(mi^2)	(cfs)	(cfs)	(cfs)	(cfs)	(cfs)
14034500	96	23,800	4,877	4,813	7,120	8,140

SUMMARY AND DISCUSSIONS

A regional regression analysis using the hybrid method was conducted to develop regression equations for Willow Creek and its tributaries in Morrow County, Oregon. Compared to conventionally developed regression equations, the hybrid regression equations use all peak records at gaging stations and a number of historic thunderstorm discharges at miscellaneous sites throughout a region. The hybrid regression equation is therefore able to reflect the impact of thunderstorms on flood frequency flows, which is a very important watershed characteristic for streams in north central Oregon. The proposed 1% annual flood discharges were compared to other hydrologic data and were considered reasonable and, therefore, recommended for use in the Morrow County Flood Insurance Study. The regression equations developed in this study are expected to be useful for other applications in the region. The hybrid regression method is a good approach when there are many historic peaks at ungaged sites and/or many gaging stations for which a conventional flood frequency analysis is not appropriate.

REFERENCES

Cunnane, C. (1978). "Unbiased plotting-positions – a review," Journal of Hydrology, 37, pp 205-222.

Fuller, W. E. (1914). "Flood flows." Trans., ASCE, 77(1293), pp 564-617.

Harris D. D., and Hubbard L. E. (1983). Magnitude and Frequency of Floods in Eastern Oregon. Water-Resources Investigations Report 82-4078, U.S. Geological Survey, Portland, Oregon.

Hjalmarson, H. W., and Thomas, B. E. (1992). "New look at regional flood-frequency relations for arid lands," Journal of Hydraulic Engineering, ASCE, 118(6), pp 868-886.

Sumioka, S. S., Kresch, D. L., and Kasnick, K. D. (1998). Magnitude and Frequency of Floods in Washington. Water-Resources Investigations Report 97-4277, U.S. Geological Survey, Tacoma, Washington.